Project AAIMS Technical Reports

Project AAIMS

2004

Reliability and Criterion Validity for Three Potential Algebra Measures

Anne Foegen
Iowa State University, afoegen@iastate.edu

Linda S. Lind
Iowa State University, afoegen@iastate.edu

Reliability and Criterion Validity for Three Potential Algebra Measures

Abstract
This technical report summarizes the results of a study in which we examined the technical adequacy of three potential algebra measures for progress monitoring. One hundred thirteen students (18 of whom were receiving special education services) completed two forms of a Basic Skills probe, an Algebra Concepts probe, and a Content Analysis probe. In addition, we gathered data on criterion variables including grades, classroom assessment records, teacher ratings, and standardized test scores. We used multiple timing periods for the Basic Skills and Algebra Concepts probes to examine the efficacy of differing durations. We examined both test-retest and alternate form reliability for individual probes of all three types and for aggregated scores from two of the probe types. Criterion validity was examined using correlations between students' probe scores and their scores from other indicators of algebra proficiency. The results of the study indicate that the Algebra Concepts probe is the most promising of the three measures investigated. It has adequate reliability and demonstrated the strongest correlations with the criterion measures. The Basic Skills probe had lower levels of reliability and more limited relations to the criterion measures (with the exception of the computation subtests of the standardized achievement tests). The Content Analysis probe had the highest levels of reliability among the three probes and moderate to moderately high correlations with several of the criterion measures. Concerns were identified about the difficulty of this probe because a large proportion of the students had scores of 10 or fewer points on the probe. Further research is needed to investigate more appropriate timing duration for the Basic Skills and Algebra Concepts probes. In the current study, the duration of the former was too short, while the duration of the latter was too long. The study should be replicated with additional, and more diverse student populations to determine the generalizability of the findings. Finally, subsequent research should examine the effects of routine progress monitoring on the measures' stability and sensitivity to growth.

Disciplines
Curriculum and Instruction | Higher Education Administration | Science and Mathematics Education

Comments
Project AAIMS is funded by the U.S. Department of Education, Office of Special Education Programs, Grant # H324C030060; Technical Report #2.

PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS

Reliability and Criterion Validity for
Three Potential Algebra Measures

Technical Report #2

Anne Foegen, Ph.D.
Linda Lind, Ph.D.

Iowa State University

September 2004

Project AAIMS is funded by the U.S. Department of Education,
Office of Special Education Programs, Grant # H324C030060

PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS

Reliability and Criterion Validity for Three Potential Algebra Measures

Technical Report 2

Abstract

This technical report summarizes the results of a study in which we examined the technical adequacy of three potential algebra measures for progress monitoring. One hundred thirteen students (18 of whom were receiving special education services) completed two forms of a Basic Skills probe, an Algebra Concepts probe, and a Content Analysis probe. In addition, we gathered data on criterion variables including grades, classroom assessment records, teacher ratings, and standardized test scores. We used multiple timing periods for the Basic Skills and Algebra Concepts probes to examine the efficacy of differing durations. We examined both test-retest and alternate form reliability for individual probes of all three types and for aggregated scores from two of the probe types. Criterion validity was examined using correlations between students' probe scores and their scores from other indicators of algebra proficiency.

The results of the study indicate that the Algebra Concepts probe is the most promising of the three measures investigated. It has adequate reliability and demonstrated the strongest correlations with the criterion measures. The Basic Skills probe had lower levels of reliability and more limited relations to the criterion measures (with the exception of the computation subtests of the standardized achievement tests). The Content Analysis probe had the highest levels of reliability among the three probes and moderate to moderately high correlations with several of the criterion measures. Concerns were identified about the difficulty of this probe because a large proportion of the students had scores of 10 or fewer points on the probe.

Further research is needed to investigate more appropriate timing durations for the Basic Skills and Algebra Concepts probes. In the current study, the duration of the former was too short, while the duration of the latter was too long. The study should be replicated with additional, and more diverse student populations to determine the generalizability of the findings. Finally, subsequent research should examine the effects of routine progress monitoring on the measures' stability and sensitivity to growth.

Full Report

Introduction

Algebra often functions in the role of a 'gatekeeper,' with proficiency in algebra having significant influence on individuals' access to higher education and professional career paths. If students with disabilities are to have access to these opportunities, it is critical that they develop proficiency in algebra. Robert Moses, a mathematics educator and civil rights advocate, sees algebra as the 'civil right' of the 21st century. He argues that algebra proficiency provides the same access to economic and social equity that the right to vote represented during the Civil Rights movement of the 1960s (Moses & Cobb, 2002). Project AAIMS (Algebra Assessment and Instruction—Meeting Standards) strives to improve student learning in algebra for all students, including those with and without disabilities. Project AAIMS has two primary objectives. First, we will examine the alignment between algebra curriculum, instruction, and assessment for students with and without disabilities. Second, we will develop and validate progress monitoring tools to support teachers' instructional decision making relative to student learning in algebra. This report describes a study of the technical adequacy of three potential measures for monitoring progress in algebra.

Method

The study described in this report was conducted in May 2004 in District A. This district serves four small towns as well as the rural agricultural areas between the towns. Approximately 7,000 residents reside in the school district. The junior/senior high school has an enrollment of approximately 600 students; about 12 percent of these students receive special education services. Approximately 13 percent of the district's students are eligible for free and reduced lunch; three percent are of diverse backgrounds in terms of race, culture and ethnicity. Data for the study were gathered on three consecutive Tuesdays in May 2004. The two rounds of algebra probe data collection occurred on the first two Tuesdays and the algebra criterion measure was administered on the third Tuesday. All data collection activities involving students were completed during regular class time. Project AAIMS staff administered all measures.

<u>Participants</u>

One hundred thirteen students in District A participated in the study. Written parental/guardian consent and written student assent were obtained for all of these students using procedures approved by Iowa State University's Human Subjects Review Committee. A description of the participating students is provided in Table 1.

Table 1. Demographic Characteristics of Student Participants by Grade Level

	Total	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
N	113	15	63	26	8	1
Gender						
Male	50	7	25	13	4	1
Female	63	8	38	13	4	0
Ethnicity						
White	110	15	60	26	8	1
Asian	2		2			
Hispanic	1		1			
Lunch						
Free	10		4	4	2	
Reduced	8	1	1	4	1	1
Disability						
IEP	18		8	7	3	

As the data in Table 1 indicate, the vast majority of the participants (96%) were white and 56% were in ninth grade, the traditional grade in which students in District A complete algebra. Sixteen percent participated in federal free and reduced lunch programs and 16% were students with disabilities who were receiving special education services. Fifteen of the students were participating in an 8th grade algebra course available to advanced students. Six of the students (all of whom had disabilities) were enrolled in algebra courses taught by special education teachers (2 in a pre-algebra course, 4 in an Algebra I course). Of the remaining students, 49 were participating in traditional Algebra I courses, and 43 were enrolled in either Algebra 1A or Algebra 1B (a pair of courses in which the traditional Algebra 1 content is taught over the course of two years, with the first half of the content presented in the first year and the second half in the second year).

Additional Information on Students with Disabilities. Because the applicability of the algebra probes to students with disabilities is an important part of Project AAIMS, additional information about the eighteen students with disabilities participating in the project is provided in Table 2.

Table 2. Descriptive Information on the Programs of Students with Disabilities

Characteristic	Quantification
% time in general education	Range = 25 –100%; Mean = 74% 50% of students spend more than 87% of their instructional time in general education
# of students with math goals	12
# of students receiving math instruction in general education classes	12
# of students receiving math instruction in a special education setting	6
# of students receiving English in a special education setting	4
# of students receiving social studies in a special education setting	3
# of students receiving science in a special education setting	3
# of students receiving health in a special education setting	1
# of students with one period of resource study hall daily	16
# of students with two periods of resource study hall daily	4
# of students with goal code D2: *Is responsible for self*	2
# of students with goal code E1: *Complies with school and community rules*	1
# of students with goal code F2: *Demonstrates competence in basic reading skills*	2
# of students with goal code F2C: *Comprehension*	11
# of students with goal code F2F*:* *Fluency*	1
# of students with goal code F3: *Demonstrates competence in basic math skills*	1
# of students with goal code F3A*:* *Applied math*	7
# of students with goal code F3C*:* *Computation*	3
# of students with goal code F4: *Demonstrates competence in basic written language skills*	3

# of students with goal code F4C:	*Composition*	5
# of students with goal code F4M:	*Mechanics of writing: punctuation, grammar, spelling*	1
# of students with goal code F5:	*Demonstrates competence in other academics*	1

Students with disabilities earned a mean GPA for the 2003-04 school year of 2.19 (range 0.80 – 3.28). In algebra, students with disabilities earned mean grades of 2.7 [B-] (range .67 [D-] to 4.0 [A]). Standardized test data for these students reflect their difficulties with academic content; in District A, the Iowa Tests of Educational Development are used as a district-wide assessment. On average, students with disabilities obtained national percentile rank scores of 39 and 34 in Concepts/Problem Solving, and Computation, respectively. They demonstrated greater deficits in reading, with mean percentile ranks of 28, 27, and 25 in Vocabulary, Reading Comprehension, and Reading Total, respectively.

Measures

Two groups of measures were used in this study. The first group consists of the curriculum-based measures of algebra performance developed by the Project AAIMS research team. The second group consists of the measures that served as criterion indicators of students' proficiency in algebra. Each of the groups of measures is described below.

Algebra Progress Monitoring Measures. Three algebra measures were examined in this study; sample copies of each are provided in the Appendix. The first, which we refer to as the *Basic Skills* measure, was designed to assess the 'tool skills' that we anticipated students would need to be proficient in algebra. Just as elementary students' proficiency with basic facts is associated with their ease in solving more complex problems, we hypothesized that there might be some basic skills in algebra that serve as indicators of overall proficiency. The tasks included in the Basic Skills measure included solving simple equations (i.e., $4 + f = 7$. $f =$ ___), applying the distributive property (i.e., Simplify: $4(3 + s) – 7$), working with integers (i.e., Simplify: $(-5) + (-4) – 1$), and combining like terms (i.e., $y^2 + y – 4y + 3y^2$). In our discussions with teachers, they frequently commented that many students had difficulty with integers and with applying the distributive property. The Basic Skills probe includes many skills one would assume that students proficient in algebra would be able to complete with reasonable levels of automaticity.

We created two parallel forms of the Basic Skills probes. Each probe consisted of 60 items; each item was scored as one point if it was answered correctly. Students marked their papers at the end of 1 minute and 2 minutes of work time, and were asked to stop working after 3 minutes. Because of the elementary level of the tasks on this probe, we anticipated that 3 minutes would be a sufficient amount of time for students to work on this task.

The second algebra progress monitoring measure that we developed was the *Algebra Concepts* measure. This task was designed to reflect five core concepts that we derived from our reading of the literature and our conversations with colleagues in mathematics education. The five concepts included (1) understanding variables and expressions (i.e., *Write the expression for this phrase: 6 less than a number*); (2) manipulating expressions involving integers, exponents, and order of operations (i.e., $2^3 =$ ___; $8 \div 2 + 4 \bullet 3 =$ ___); (3) basic graphing (i.e., graphing inequalities on number lines and identifying y-intercepts and slopes from coordinate graphs); (4) solving simple equations (i.e., Evaluate $2x + y$ when $x = 2$ and y=3); and (5) tasks involving patterns and functions (i.e., complete the missing element in a table that includes a function and all the entries but one). Our intent with this measure was to assess the extent to which students were proficient in solving problems that addressed these core concepts of early algebra. It is important to note that with this task, many of the items represented concepts and skills that would be learned as part of pre-algebra or very early instruction in an Algebra I course, if not earlier. We recognized that proficiency on this task was not equivalent to having mastered all the concepts taught in Algebra I, but we hoped to determine whether proficiency on these basic skills might serve as an indicator of more general proficiency in algebra.

The Algebra Concepts probe consisted of 42 items; eight of these items required two responses, so 50 total points were possible on this probe. We created two parallel forms of this probe. As with the Basic Skills probe, we asked students to mark their progress at the end of five and six minutes of work time. Students were asked to stop working after seven minutes.

The third task examined in the study was the *Content Analysis* probe. We developed two parallel forms of this task by analyzing the content taught in the algebra textbook. Because all three districts participating in Project AAIMS are using the same textbook series, we wanted to investigate a measure that was directly derived from the instructional materials. We developed the items by sampling from the chapter tests and reviews. We sought to identify items that represented core concepts/problem types in each chapter. Based on teacher feedback, we sampled chapters in the middle portion of the text at a higher rate (two questions per chapter) than the chapters at the beginning (review) and end (advanced concepts/skills) sections of the text. We anticipated that this probe might provide a more direct reflection of the extent to which students had learned the content of instruction than would the other probes, which represented more general indicators of algebra proficiency.

The Content Analysis probes consisted of 16 items. Students had ten minutes to work on the probe and were instructed to attempt all of the problems. The number of points possible on the probes was 53 for Form 1 and 59 for Form 2. The number of points possible for the individual problems ranged from one to six. Because this probe required more complex problem solving and fewer responses, we needed a means of increasing the sensitivity of the scores. We asked the algebra teachers to solve each of the problems, showing the steps and the final solution in the same way that they would expect students to show their work. We then assigned a point to each step of the solution. Students were awarded points corresponding to any of the steps they completed correctly in their responses. In the directions for this probe, we encouraged students to show their work if necessary to obtain 'partial credit' even if they weren't able to solve the entire problem. We also informed them that if they were able to complete the problems without showing all the steps, they would be awarded the full number of points possible for the correct solution. We opted to use this practice in order to reinforce/reward students who were so proficient that it would be tedious for them to record each step of the problem.

<u>Criterion Measures</u>. In order to evaluate the criterion validity of the three algebra progress monitoring measures, we gathered data on a variety of other indicators of students' proficiency in algebra. Some of these measures were based on students' performance in class (and in school more generally) and their teachers' evaluation of their proficiency. Other measures reflected students' performance on standardized assessment instruments.

The classroom-based measures included grade-based measures, classroom performance measures, and teacher ratings. Each student's *algebra grade*, the grade s/he earned in algebra during the spring semester of the 2003-04 school year, was recorded using a four-point scale (i.e., A = 4.0, B = 3.0). *GPA* represented students' overall grade point average for the 2003-04 academic year and was recorded using the same four-point scale, with scores rounded to the nearest hundredth. The *homework completion* measure represented the percent of homework assignments that students had completed. The *homework accuracy* measure represented the average percent correct score earned on homework assignments. The *test/quiz accuracy* measure reflected students' average percent correct on tests and quizzes. All three of these measures were calculated using grade book records provided by the participating teachers.

We also wanted to include the teachers' evaluations of students' proficiency in algebra. To accomplish this, we asked each teacher to complete a *teacher rating* form for all the students to whom s/he taught algebra. Student names were alphabetized across classes to minimize any biases that might be associated with particular class sections. Teachers used a 5-point Likert scale (1=low proficiency, 5= high proficiency) to rate each student's proficiency in algebra in comparison to same-grade peers.

Student performance on standardized, norm-referenced assessments was evaluated using school records and with an algebra instrument administered as part of the project. In District A, 8^{th} grade students complete the *Iowa Tests of Basic Skills* (ITBS) each spring. Students in grades 9 through 11 complete the *Iowa Tests of Educational Development* (ITED). District records were used to access students' scores on these instruments; national percentile ranks were used for the analyses. For the ITBS, the following scores were recorded: Problems/Data, Concepts/Estimation, Computation, Math Total, Reading Total. For the ITED, we recorded the Concepts/Problems score (which was identical to the Math Total score), the Computation score, and the Reading Total score.

Because neither of the district-administered measures provided a direct assessment of algebra, we also administered the *Iowa Algebra Aptitude Test* (IAAT). This norm-referenced instrument is typically used to evaluate the potential of 7^{th} grade students for successful study of algebra in 8^{th} grade. Although we recognized the limitations of using this aptitude measure, we were unable to identify a norm-referenced test of algebra achievement.

Procedures

The algebra probes were administered in a single 45-minute class period. During each class, students completed two parallel forms of the Basic Skills probe, two parallel forms of the Algebra Concepts probe, and one of two parallel forms of the Content Analysis probe. The order in which the Basic Skills and Algebra Concepts probes were administered was counterbalanced across classes, as was the order of each of the parallel forms. Students always completed the Content Analysis probe last. Table 3 depicts the order in which the probes were administered during each of the two testing sessions.

During the administration of the Basic Skills probes and the Algebra Concepts probes, students were instructed to mark their papers when the administrator said the word "slash" during the administration period. This was done to investigate the appropriateness of several different timing periods for each probe. A copy of the standardized directions used for each administration session is provided in the Appendix.

Table 3. Administration Schedule for Probe Forms by Period

Session	8^{th} grade algebra	Algebra 1A	Algebra 1B	Algebra (Per. 3)	Algebra (Per. 4)	Algebra (Per. 5)	SpEd Algebra	SpEd Algebra
1								
	A2	A2	A1	B2	A1	B1	A2	B1
	A1	A1	A2	B1	A2	B2	A1	B2
	B1	B1	B1	A1	B2	A2	B2	A2
	B2	B2	B2	A2	B1	A1	B1	A1
	C2	C1	*	C2	C1	C2	C1	C2
2								
	A2	A2	A1	B2	A1	B1	A2	B1
	A1	A1	A2	B1	A2	B2	A1	B2
	B1	B1	B1	A1	B2	A2	B2	A2
	B2	B2	B2	A2	B1	A1	B1	A1
	C1	C1	C2	C2	C2	C2	C1	C1

A1, A2 = Basic Skills probes 1 and 2
B1, B2 = Algebra Concepts probes 1 and 2
C1, C2 = Content Analysis probes 1 and 2
*Note. Because of an error in photocopying, students in this class period received a copy of the C1 probe that contained two copies of the first page of the probe. Because of the potential that this error may have limited their scores on the probes, no scores were recorded for this class on this probe.

Results

Scoring Reliability

Scoring accuracy was evaluated by re-scoring a portion of the probes. For each probe, an answer-by-answer comparison was conducted and an interscorer reliability estimate was calculated by dividing the number of agreements by the total number of answers scored. These individual probe agreement percentages were then averaged across all the selected probes of a common type to determine an overall average.

We selected the probes to be re-scored by drawing from each of the class periods across the two administration periods. The two special education classes were omitted because of small student numbers (2 and 4 students in each class, respectively). Each form of the Basic Skills probes was rescored for 4 of the 16 class periods (25%), as were each form of the Algebra Concepts probes. For the Content Analysis probes, Form 1 was rescored for 3 of the 8 classes in which it was administered (38%). Form 2 was rescored in 2 of the 8 classes (25%). The number of student papers rescored and the average agreement for each form of the probe are reported in Table 4 below.

The Basic Skills and Algebra Concepts probes were scored with high levels of accuracy. The Content Analysis probes were clearly more difficulty to score consistently, although the scoring accuracy for Form 2 exceeded the minimum level for acceptable agreement that we had established at 90%. In future scoring efforts, a more detailed scoring key for the Content Analysis probes will be developed to further improve the consistency with which the probes are scored.

Table 4. Interscorer Agreement Rates and Student Papers Rescored

Probe	# Papers Rescored	Range of Agreement	Mean % Agreement
Basic Skills, Form 1	68	83 – 100%	98
Basic Skills, Form 2	60	89 – 100%	99
Algebra Concepts, Form 1	71	88 – 100%	98
Algebra Concepts, Form 2	59	88 – 100%	98
Content Analysis, Form 1	48	64 – 100%	88
Content Analysis, Form 2	35	60 – 100%	91

<u>Descriptive Data on Score Ranges and Distributions</u>

Table 5 lists the ranges, means, and standard deviations for each of the forms of the three probes. The results for the Basic Skills probe indicate that there may be somewhat of a floor effect for a 1-minute timing. Mean scores for the 1-minute timings were all less than nine problems correct. A review of the frequencies revealed that the percentage of the sample having five or fewer problems correct in one minute was 26, 30, 46, and 60 in each of the four administrations of a Basic Skills probe for one minute. This effect is likely to limit the utility of this measure for monitoring the progress of low-performing students when a 1-minute administration is used. Given that the maximum score for this probe type is 60, there is little evidence that a ceiling effect will pose any difficulties for monitoring the progress of students about whom there is concern regarding their algebra performance. In the future, administration times for this probe should focus on two-, three-, and four-minute periods.

Results for the Algebra Concepts probe suggest that there may be ceiling effects in the time limits examined. The maximum score for this probe is 50, which was achieved by some students in the 5-minute administration. A review of the frequencies of scores for the 5-minute administration revealed that if a score of 35 was used as a threshold (allowing 15 points for growth), the percentage of each sample that obtained scores at or above 35 was 7, 8, 29, and 31. The latter two figures were obtained the second week of administration, suggesting that familiarity with the probe format increases students' rates of correct responding. In the future, administration times for this probe should focus on periods of two, three, and four minutes.

Results for the Content Analysis probe suggest possible problems with the difficulty level of the probe. The maximum number of points possible on the probe was 53 for form C1 and 59 for form C2. On C1, 32% of students who completed this probe the first data collection session earned scores of ten or less; 34% of students who completed this probe on the second data collection session earned scores of ten or less. For the C2 form, 36% of students who took the probe the first data collection session earned scores from zero to ten and 44% of students in the second data collection session earned scores from zero to ten on this probe. These problems are magnified if the results for the 8th grade (advanced) class are considered separately. While students in the 8th grade class were able to complete the probe with good success (scores from 29-41 and 23-40 for each administration), they account for the bulk of the scores over 20 in the distribution of all students. These results, especially given the fact that the probe was

Table 5. Descriptive Data for Three Algebra Probes Across Administration Sessions

Measure	Session	Timing in minutes	N	Range	Mean	Standard Deviation
Basic Skills Form 1	1	1	105	1 – 16	6.39	3.33
		2	105	3 – 33	11.70	6.05
		3	105	3 – 41	16.71	8.07
	2	1	104	1 – 27	8.51	4.34
		2	104	3 – 44	15.59	7.21
		3	104	5 – 44	21.88	8.96
Basic Skills Form 2	1	1	108	0 – 14	5.26	2.97
		2	108	2 – 25	10.38	5.32
		3	108	4 – 35	15.09	7.31
	2	1	103	2 – 21	8.16	4.29
		2	103	3 – 36	14.60	7.09
		3	103	7 – 48	20.16	9.09
Algebra Concepts Form 1	1	5	109	2 – 50	20.83	9.58
		6	109	3 – 50	24.37	10.57
		7	109	3 – 50	27.20	10.86
	2	5	104	2 – 50	27.95	11.58
		6	104	2 – 50	30.89	11.34
		7	104	2 – 50	32.04	11.13
Algebra Concepts Form 2	1	5	106	1 – 45	22.43	8.98
		6	106	1 – 49	26.65	10.32
		7	106	1 – 50	28.97	10.93
	2	5	107	3 – 50	27.62	11.94
		6	107	3 – 50	30.61	11.70
		7	107	3 – 50	31.80	11.81
Content Analysis Form 1	1	-	31	0 – 26	11.35	7.73
	2	-	38	0 – 41	19.68	13.96
Content Analysis Form 2	1	-	55	1 – 40	18.33	11.17
	2	-	66	4 – 31	13.27	6.45

administered at the conclusion of a year of algebra instruction, suggest that the probe (in its current form) is not likely to be useful in monitoring student progress over the course of a year. If 30 to 45% of students earn scores of ten or less after a year of instruction, it is likely that there will be serious floor effects if this probe is to be used in the fall. This issue will be tested empirically when this study is replicated with additional districts in Fall 2004.

Reliability of Individual Probe Scores

The reliability of individual probes was evaluated by examining alternate form reliability (the correlation between the two forms of the A and B probes given during the same data collection session) and test-retest reliability (the correlation between the same form of each probe given across the two data collection sessions). The results of these analyses for the Basic Skills probes are presented in Table 6.

Table 6: Reliability results for single basic skills probes

Probe Type		Alternate Forms		Test-Retest
Basic Skills-1 min.				
First session		.68	Form 1	.68
Second session		.78	Form 2	.53
Basic Skills – 2 min.				
First session		.66	Form 1	.80
Second session		.76	Form 2	.68
Basic Skills – 3 min.				
First session		.71	Form 1	.84
Second session		.86	Form 2	.81

Note: All correlations significant at $p < .05$.

The results in Table 6 corroborate the concerns about possible floor effects identified when the distributions for the Basic Skills probes were examined. The reliability of the basic skills probes ranged from .53 to .78 for the 1-minute period. In general, reliability increased as the length of time students were allowed to complete the probe increased. In addition, the reliability increased from the first to the second session of data collection at all three timing lengths. This result suggests that as students become more familiar with the task, their performance becomes more stable. The 3-minute timing produced the best results for a single administration of the Basic Skills probe. Across all three timing levels, Form 1 proved to be more reliable than Form 2, despite the fact that the forms had been counterbalanced across students. While these differences diminished at the 3-minute level, it is unclear why a parallel form would produce such different reliability results.

The results for the Algebra Concepts probe, presented below in Table 7, parallel the findings obtained when the distributions of scores were examined. The limited increases in reliability with increased time are consistent with possible ceiling effects. The reliability of the Algebra Concepts probe is fairly strong at five minutes with a single administration and meets expected reliability levels in three of four instances at six minutes. As with the Basic Skills probe, there seems to be a slight ‘learning curve’ that occurs in becoming familiar with the probe format and contents. At all three time levels, students’ scores on the second day of data collection were more consistent than on the first. There were some differences in the reliability of the two forms of the probe on the 5-minute timing, but these diminished when the data for the 6- and 7-minute timings were considered.

Table 7: Reliability results for single algebra concepts probes

Probe Type		Alternate Forms		Test-Retest
Algebra Concepts-5 min.				
First session		.72	Form 1	.86
Second session		.82	Form 2	.75
Algebra Concepts-6 min.				
First session		.73	Form 1	.86
Second session		.92	Form 2	.83
Algebra Concepts-7 min.				
First session		.82	Form 1	.88
Second session		.94	Form 2	.86

Note: All correlations significant at $p < .05$.

Results for the Content Analysis probe are presented below in Table 8. Of the three alternatives explored in this study, the Content Analysis probe had the strongest levels of reliability; this may be due in part to the extended time available for this probe in comparison to the other formats. Unfortunately, students' scores on the Content Analysis probe were extremely low, limiting the utility of this probe format for progress monitoring.

Table 8: Reliability results for the content analysis probes

Probe Type		Alternate Forms		Test-Retest
Content Analysis				
Across Sessions		.89	Form 1	.85
			Form 2	.86

Note: All correlations significant at $p < .05$.

Reliability of Aggregated Probe Scores

Because students completed two forms of the Basic Skills and Algebra Concepts probes during each data collection session, it was also possible to examine the effects of aggregating scores from two probes on the resulting reliability levels. Previous research in other areas of mathematics (Foegen, 2000; Fuchs, Deno, & Marston, 1983) has determined that for some types of mathematics skills and concepts, multiple probes need to be aggregated to obtain reliable scores for individual students. Table 9 presents the results for the aggregated scores on the Basic Skills and Algebra Concepts probes. The alternate form coefficients were computed by averaging scores from the two administrations of Form 1 and repeating this process for each form and each timing variation. The test-retest coefficients were computed by averaging scores from the two forms of each probe administered on the first data collection day, and then correlating these scores with the averaged scores for the same probes from the second data collection day.

The results in Table 9 indicate that aggregation of two probe scores did not result in substantial improvements in the reliability of the Basic Skills probe for alternate form reliability or for the 1- and 2 minute time limits for test-retest reliability. For the 3-minute probe, the coefficient obtained for two aggregated scores was roughly equivalent to the coefficients obtained for single scores (.86 vs. .84 and .81). These results lead to the conclusion that Basic Skills probes are sufficiently reliable when single probes are used. Only minimal gains are achieved by aggregating two scores. Future research should examine the extent to which the scores of single probes become increasingly stable with repeated administration, thus eliminating any need for considering aggregation.

Table 9. Reliability for Aggregated Basic Skills and Algebra Concepts Probes

Probe	Alternate Form Reliability	Test-Retest Reliability
Basic Skills		
1 minute	.67	.63
2 minute	.75	.78
3 minute	.83	.86
Algebra Concepts		
5 minutes	.85	.88
6 minutes	.88	.91
7 minutes	.91	.91

Note: All correlations significant at $p < .05$.

With regard to the Algebra Concepts probe, aggregation at five minutes produced higher reliability coefficients than did any of the single 5-minute administrations. This finding was consistent across both test-retest and alternate form reliability results. At six and seven minutes, the coefficients for aggregated probes were at the high end of the range produced by single probes. Given that single administrations at six and seven minutes proved to be highly reliable, the slight gains in the coefficients obtained with aggregation do not seem to merit the ‘costs’ of doubling the administration time.

Criterion Validity for Individual Probes

The criterion validity of the measures was examined by correlating scores on the probes with the criterion measures that served as additional indicators of students’ proficiency in algebra. The indicators we used included students’ overall grade point average (GPA) and grades in algebra; teachers’ evaluations of student proficiency; scores on homework, tests, and quizzes; scores from standardized tests in mathematics administered by the district; and scores obtained from a norm-referenced test of algebra aptitude, the Iowa Algebra Aptitude Test (IAAT). In the following section, the correlation coefficients between scores on the three algebra measures and each of these criterion variables is presented and discussed. Correlation coefficients for each of the probes with the respective criterion measures are presented in Table 10. Because a number of correlation coefficients were produced in the analyses (scores from each of two forms of the Basic Skills and Algebra Concepts probes were available for each of the two administration days), mean correlations are reported. The range of obtained correlations is included in parentheses. If at least two of the four correlations were significant, the mean correlation is reported.

Table 10. Criterion Validity Results for Single Probes: Mean Correlation Coefficients and Ranges

	Algebra Probes						
Criterion Measure	Basic Skills			Algebra Concepts			Content Analysis
	1 min	2 min	3 min	5 min	6 min	7 min	10 min.
Overall GPA	.43** (.34 - .57)	.46.** (.42 - .57)	.49** (.42 - .54)	.58** (.47 - .66)	.62** (.53 - .67)	.63** (.55 - .68)	.63** (.38 - .89)
Grade in Algebra	NS[a] (3/4[b]; .30**)	.23* (1 NS)	.26* (.21 - .30)	.32** (.23 - .37)	.35** (.24 - .39)	.35** (.24 - .38)	.43** (.33 - .65)
Teacher Rating	.40**(.32 - .50)	.43** (.36 - .50)	.45** (.43 - .49)	.66** (.53 - .73)	.69** (.56 - .75)	.70** (.59 - .76)	.74** (.52 - .92)
Class Data							
Hw completion	.21* (1 NS, .20 - .22)	.23* (1 NS, .20 - .25)	.24* (2 NS, .21 - .26)	.23* (.20 - .25)	.26** (.24 - .27)	.27* (.22 - .28)	NS
Hw accuracy	.26** (1 NS, .23 - .29)	.28** (1 NS, .24 - .32)	.28** (1 NS, .22 - .33)	.29** (.27 - .31)	.32** (.31 - .34)	.32** (.28 - .36)	.37* (1 NS, .31 - .41)
Test/quiz acc.	NS	.28** (1 NS, .20 - .33)	.28** (.23 - .35)	.35** (.32 - .40)	.37** (.34 - .39)	.37** (.34 - .39)	.47** (1 NS, .35 - .53)
ITBS Scores[c]							
Math Total	NS	NS	NS	.69* (2 NS, .63 - .74)	.68* (2 NS, .56 - .80)	.83** (2 NS, .81 - .85)	NS
Prob/Data	NS	NS	NS	.65* (2 NS, .57 - .72)	NS (.78**, 3 NS)	.85** (2 NS, .84 - .86)	NS
Concepts/Est	NS	NS	NS	.67* (2 NS, .62 - .72)	.70* (2 NS, .62 - .77)	.68* (2 NS, .62 - .73)	NS
Computation	.67* (1 NS, .60 - .70)	.73** (1 NS, .7 - .75)	.73** (.63 - .77)	.62* (2 NS, .61 - .62)	.67* (1 NS, .58 - .85)	.67* (1 NS, .60 - .76)	NS
Reading Total	NS	NS	NS	NS	NS	NS	NS

[a] NS = nonsignificant
[b] Three of four obtained correlations were nonsignificant
[c] Only 8th grade students completed the ITBS; all other students completed the ITED. Therefore, ITBS scores are based on Ns of 14 to 15

Table 10, continued.

	Algebra Probes						
Criterion Measure	Algebra Basic Skills			Algebra Concepts			Content Analysis
	1 min	2 min	3 min	5 min	6 min	7 min	10 min.
ITED Scores							
Con/Prob (aka Math Total)	.34** (.31 - .47)	.42** (.37 - .49)	.44** (.39 - .54)	.60** (.57 - .64)	.64** (.61 - .66)	.65** (.63 - .66)	.50** (1 NS, .39 - .56)
Computation	.34 **(.30 - .40)	.34** (.30 - .39)	.35** (.30 - .44)	.65** (.59 - .71)	.70** (.68 - .71)	.72** (.71 - .72)	.54** (1 NS, .48 - .57)
Reading Total	.32** (1 NS, .28 - .34)	.31** (.26 - .36)	.30** (.24 - .35)	.46** (.42 - .54)	.46** (.42 - .49)	.48** (.43 - .50)	.37* (2 NS, .29 - .44)
IAAT Scores							
Total	.51** (.47 - .55)	.51** (.47 - .55)	.52** (.50 - .53)	.69** (.55 - .75)	.73** (.62 - .80)	.75** (.67 - .80)	.73** (.61 - .94)
Part A	.42** (.38 - .45)	.42** (.38 - .45)	.43** (.40 - .50)	.62** (.50 - .66)	.65** (.55 - .71)	.67** (.60 - .72)	.67** (.48 - .90)
Part B	.48** (.44 - .50)	.50** (.47 - .54)	.52** (.51 - .53)	.65** (.55 - .70)	.68** (.59 - .73)	.70** (.64 - .74)	.67** (.59 - .83)
Part C	.47** (.43 - .53)	.48** (.40 - .52)	.48** (.43 - .51)	.65** (.51 - .72)	.71** (.59 - .79)	.73** (.65 - .79)	.72** (.60 - .90)
Part D	.48** (.42 - .55)	.49** (.44 - .51)	.49** (.47 - .51)	.64** (.50 - .73)	.68** (.57 - .74)	.69** (.62 - .72)	.66** (.52 - .91)

Correlations with the *grade-based measures* revealed that the overall GPAs were more strongly related to performance on the algebra probes than were students' grades in algebra. This is not surprising, given that the overall GPA represents a composite of academic performance. What was surprising was the strength of the correlations with the Algebra

Concepts and Content Analysis probes. Based on our past research experiences (Foegen, 2000; Foegen & Deno, 2001), correlations between progress monitoring measures and grade-based measures are often low at best (often in the .3 to .4 range) and frequently non-significant, in part because grades include much more than isolated academic achievement. Students' work habits, motivation, and attitude also influence the grade a teacher assigns. We were surprised to see that two of the algebra probes had correlations in the .6 range with overall GPA.

Scores obtained from the *teacher rating* of algebra proficiency revealed low correlations with the Basic Skills probe and moderate to strong correlations with the Algebra Concepts and Content Analysis probes. For the two probes that had variations in timing, only minor increases (.03 - .05) in criterion validity were obtained as the duration of the probes increased.

Class data included the percentage of homework completed, the average percent correct on homework assignments, and the average percent correct on quizzes and tests. These criterion variables showed limited relations with the algebra probes, with the majority of the correlations in the .2 to .3 range. Of the three types of probes, the strongest relation was identified between the Content Analysis format and percentage correct on tests/quizzes (.47).

Two types of *standardized achievement test* data were included in the analysis. Students in the eighth grade algebra class complete the ITBS, so the data in the table's ITBS section reflect only students in the eighth grade algebra class. For the Basic Skills probe, only the Computation subtest was found to relate to students' probe scores. These correlations were quite strong (.67 - .73) and reflect the inclusion of a substantial number of basic computation problems (of the form $a + 5 = 12$) on this probe. Likewise, students' scores on the Content Analysis probe were not correlated with their performance on the ITBS in any of the subtests. The Algebra Concepts probe showed moderate to strong relations with student performance on the ITBS. Correlations ranged from .62 to .85, with the correlations for the Problems/Data and Concepts/Estimation subtests much stronger than those for the Computation subtest. This result is consistent with the relatively minimal emphasis on computation in this task. Readers will note that there were several non-significant findings among the results for this probe. The analyses produced an unusual pattern in which the results for probes administered on the first day of data collection produced very high correlations, while those on the second day produced non-significant results. To further explore this issue, we examined the student data and found that one student had been absent the first day and was present the second day. No other anomalies were found in the data. It may be that the small sample size (14 students for the Day 1 analyses and 15 for Day 2) influenced this pattern of results.

Finally, it is important to note that across all three probe types, there were no relationships between students' scores on the algebra probes and their performance on the reading subtest. Whenever mathematics probes introduce reading through noncomputational problems, there is potential that the probes will reflect reading ability to a greater extent than mathematics ability. This did not occur for this group of students, but it is important to note that these were advanced students who had been selected to have the opportunity to take algebra in eighth grade.

The remainder of the students in the sample completed the ITED as their district-wide achievement measure. Scores were available for two mathematics subtests: Concepts/Problems and Computation. In the district records, a Total Math score was also listed. Because this score was identical to the Concepts/Problems score in all cases, it was not included in the analyses. Reading scores were also included in the analyses to determine the extent to which reading proficiency might be associated with performance on the algebra probes. The results for the Basic Skills and Algebra Concepts probes were similar to those obtained for the ITBS scores in the eighth grade class. Correlations with the Basic Skills probes were in the low range (.30 - .44), while scores on the Algebra Concepts probe correlated with ITED performance in the moderate range (.60 - .72). Correlations between algebra probe scores and performance on the reading subtest of the ITED were in the .3 range for the Basic Skills probe and ranged from .46 - .48 for the Algebra Concepts probe. The correlations obtained for the Content Analysis probe were between those for the Basic Skills probe and the Algebra Concepts probe, and ranged from .50 to .54.

The *algebra achievement* measure consisted of four subscale scores and a total score from the IAAT. The subscales included Part A: Interpreting Mathematical Information, Part B: Translating to Symbols, Part C: Finding Relationships; and Part D: Using Symbols. Correlations between the IAAT subtest and total test scores were in the .4 to .5 range for the Basic Skills probe. For the Algebra Concepts probe, coefficients were in the .6 to .75 range. The Content Analysis coefficients were in the .66 to .73 range. Across all three probe types, the correlation coefficients for the IAAT total were equal to or exceeded those for the individual subtests. No distinctive patterns emerged for the individual subtests.

Summary of Criterion Validity Correlation Coefficients for Individual Probes

As we examined the criterion validity data across the three probes, we found that the Basic Skills probe generally had the lowest correlations, with most coefficients in the .3 to .5 range. The validity estimates fall below those obtained in existing research on progress monitoring measures for the secondary level, though all of this earlier research has been conducted at the middle school level. If this result is replicated in the two additional districts, we will conclude that the Basic Skills measure is not an effective tool for monitoring student progress in algebra. In contrast, both the Algebra Concepts probe and the Content Analysis probe had stronger criterion validity data. For the Algebra Concepts probe, the correlation coefficients were generally in the .6 to .8 range, while coefficients for the Content Analysis probe ranged from .5 to .7. Given the concerns about potential floor effects with the Content Analysis probes, we would argue that these data support the Algebra Concepts probe as the strongest candidate of the three options investigated in this study.

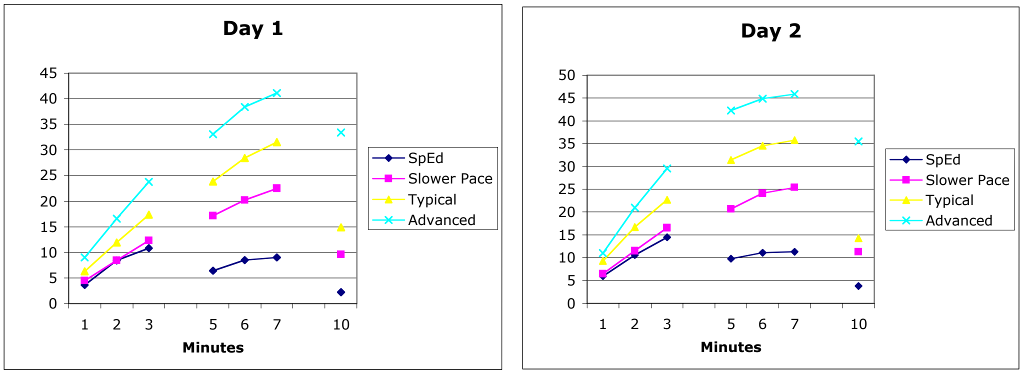
Discrimination Between Groups

As a second means of investigating the validity of the measures, we examined whether the scores of students in different algebra options differed at a level that was statistically significant. To conduct this analysis, we labeled each algebra class as belonging to one of four groups: advanced (8th grade Algebra), typical (Algebra I), and 'slower pace' (Algebra 1A, Algebra 1B), and special education (Special Education Algebra). Because the students enrolled in Algebra 1A/1B are completing the content of Algebra 1 across a two-year span, we selected the label of 'slower pace' for this group. We opted to aggregate scores from similar probes collected on the same day to minimize the number of tests required. We ran analyses of variance to determine whether scores on the probes were significantly different. In every case, the overall ANOVA was statistically significant. We then used Scheffe post-hoc multiple comparison tests to identify where significant differences between the groups were found. For the Basic Skills and Algebra Concepts probes, the patterns were consistent across the timing alternatives, so the results are discussed as a whole. Means and standard deviations for each group on each probe are listed below in Table 11. Data on the means for each of the measures are depicted in graphic form in Figure 1.

Table 11. Means and Standard Deviations on Three Probes by Group Type

Probe and Day	Advanced		Typical		Slower Pace		Special Ed.	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Basic Skill-Day 1								
1 minute	9.00	2.81	6.26	2.36	4.49	2.06	3.58	3.12
2 minutes	16.57	4.48	11.94	4.79	8.38	3.38	8.42	7.43
3 minutes	23.68	6.19	17.35	4.79	12.30	4.55	10.83	7.87
Basic Skill-Day 2								
1 minute	11.03	2.79	9.27	4.64	6.46	2.57	6.00	3.82
2 minutes	21.03	6.27	16.64	6.61	11.54	4.60	10.58	5.89
3 minutes	29.57	9.02	22.7	7.84	16.55	6.39	14.50	7.71
Algebra Concepts Day 1								
5 minutes	33.14	6.58	23.79	6.62	17.15	5.48	6.42	1.66
6 minutes	38.36	6.27	28.39	7.38	20.19	6.02	8.50	3.05
7 minutes	41.14	6.08	31.46	7.83	22.48	6.68	9.00	2.77
Algebra Concepts Day 2								
5 minutes	42.33	6.23	31.50	7.30	20.65	8.14	9.75	5.51
6 minutes	44.90	4.96	34.63	7.00	24.06	8.45	11.08	6.73
7 minutes	45.93	4.25	35.77	6.92	25.43	8.53	11.25	6.96
Content Analysis Day 1	33.36	4.72	14.85	6.91	9.58	6.55	2.17	2.32
Content Analysis Day 2	35.47	3.66	14.32	7.18	11.26	4.75	3.83	4.17

Figure 1. Mean probe performance by group status.



As illustrated in Figure 1, the data reflect the expected pattern of increasing scores across group types. The scores for 1, 2, and 3 minutes are from the Basic Skills probes, while the scores for 5, 6, and 7 minutes are for the Algebra Concepts probes and the 10 minute data reflect scores on the Content Analysis probes. The delineation between groups is most clear for the Algebra Concepts probe. This visual observation is generally supported by the results of the post hoc comparisons, with some exceptions that are likely due to small sample sizes and variability within the subgruops. On the Basic Skills probe, the slower pace, typical, and advanced groups were all significantly different from each other. The special education group was significantly different from the advanced group, but did not differ from the slower pace and typical groups. On the Content Analysis probe, the advanced group was different from all other groups, but the slower pace group could not be distinguished from the special education group or the average group. The Algebra Concepts probe was the only format in which all groups could be differentiated from each other.

Criterion Validity for Aggregated Probe Scores

In our earlier analyses, we found that only limited gains in reliability were obtained when the scores from two forms of an algebra probe were aggregated. In Table 12, we report the criterion validity coefficients using aggregated scores for each of the probes. To aggregate, we averaged the two scores of a probe type that were administered on the same day. This produced two scores for the Basic Skills and Algebra Concepts probes (Day 1 aggregate, Day 2 aggregate). Because only a single version of the Content Analysis probe was administered each testing session, our aggregate of the two scores produced a single Content Analysis score.

In reviewing the data in Table 12, we found that the general pattern was one in which aggregation had the effect of reducing the magnitude of the criterion validity correlation coefficients. There were two general exceptions to this pattern. For the Basic Skills probe, all but two of the eight coefficients computed using the ITED and IAAT scores improved for the 1-minute timing. (We arbitrarily selected 4 percentage points as denoting a substantive increase.) This result is not surprising, given the floor effects obtained for this probe. By aggregating across two samples, we likely increased the stability of the 1-minute estimate. This same increase was obtained for the 2- and 3-

Table 12. Criterion Validity Results for Aggregated Probes: Mean Correlation Coefficients and Ranges

	Algebra Probes						
Criterion Measure	Basic Skills			Algebra Concepts			Content Analysis[b]
	1 min	2 min	3 min	5 min	6 min	7 min	10 min.
Overall GPA	.39** (.34 - .43)	.41** (.39 - .42)	.42** (.39 - .45)	.54** (.53 - .54)	.54** (.53 - .55)	.55** (.54 - .56)	.61**
Grade in Algebra	.20** (1 NS, .20)	.21* (1 NS, .21)	.23* (1 NS, .23)	.25** (.25 - .25)	.26** (.25 - .27)	.27** (.25 – 28)	.35**
Teacher Rating	.41**(.33 - .48)	.42** (.40, .44)	.44** (.41 - .46)	.59** (.57 - .61)	61** (.58 - .63)	.62** (.59 - .64)	.68**
Class Data							
Hw completion	NS	NS	NS	NS	NS	NS	NS
Hw accuracy	.20* (1 NS, .20)	.20* (1 NS, .20)	.25* (1 NS, .25)	.23* (1 NS, .23)	.23* (.22 - .24)	.23* (.22 - .23)	.22*
Test/quiz acc.	NS	NS	NS	.27** (.25 - .28)	.29** (.26 - .32)	.31** (.28 - .34)	.29**
ITBS Scores[a]							
Math Total	NS	NS	NS	.74** (1 NS, .74)	.77** (1 NS, .77)	.87** (1 NS, .87)	NS
Prob/Data	NS	NS	NS	.69** (1 NS, .69)	.77** (1 NS, .77)	.89** (1 NS, .89)	NS
Concepts/Est	NS	NS	NS	.72** (1 NS, .72)	.69** (1 NS, .69)	.72** (1 NS, .72)	NS
Computation	.67**(.57 - .76)	.70** (.59 - .80)	.74** (.71 - .76)	.64** (.62 - .66)	.66** (.59 - .72)	.70** (1 NS, .70)	NS
Reading Total	NS	NS	NS	NS	NS	NS	.40**

Table 12, continued.

Criterion Measure	Algebra Probes: Basic Skills			Algebra Concepts			Content Analysis[b]
	1 min	2 min	3 min	5 min	6 min	7 min	10 min.
ITED Scores							
Con/Prob (aka Math Total)	.38** (.34 - .41)	.40** (.37 - .43)	.43** (.40 - .45)	.54** (.52 - .55)	.56** (.55 - .57)	.58** (.58 - .58)	.46**
Computation	.34** (.33 - .34)	.35** (.34 - .35)	.38** (.36 - .40)	.60** (.56 - .64)	.64** (.61 - .66)	.64** (.60 - .67)	.54**
Reading Total	.38**(.33 - .43)	.42** (.38 - .45)	.41** (.39 - .42)	.44** (.42 - .45)	.46** (.45 - .47)	.47** (.47 - .47)	.40**
IAAT Scores							
Total	.55** (.53 - .57)	.53** (.53 - .53)	.54** (.53 - .54)	.65** (.62 - .68)	.70** (.67 - .72)	.71** (.69 - .72)	.65**
Part A	.46** (.43 - .49)	.45** (.45 - .45)	.46** (.44 - .47)	.57** (.53 - .61)	.60** (.57 - .63)	.63** (.60 - .63)	.58**
Part B	.52** (.52 - .52)	.53** (.52 - .54)	.54** (.54 - .54)	.58** (.55 - .60)	.62** (.60 - .64)	.63** (.62 - .64)	.57**
Part C	.44** (.40 - .47)	.41** (.40 - .41)	.40** (.39 - .41)	.58** (.55 - .60)	.63** (.59 - .66)	.65** (.62 - .67)	.61**
Part D	.56** (.54 - .58)	.54** (.53 - .54)	.55** (.54 - .55)	.64** (.60 - .68)	.69** (.66 - .71)	.69** (.67 - .70)	.65**

[a]NOTE: Only 8th grade students completed the ITBS; all other students completed the ITED. Therefore, ITBS scores are based on Ns of 14 –15

[b]Aggregated scores for probe C were computed by averaging each student's score from the single C probes administered in each data collection session.

minute scores with the ITED reading subtest and Part D of the IAAT. We are uncertain as to why these subtests would reflect a different pattern than the others.

The second exception to the general pattern of decreasing coefficients was for the correlations between the Algebra Concepts probe and the ITBS scores (note that this is specific to the 8^{th} grade class). In eight of twelve correlations between ITBS mathematics subtests and Algebra Concepts scores, an increase resulted when scores from aggregated, rather than individual probes were used. Readers should note that the same anomaly involving non-significant correlations for the Day 2 data occurred in this analysis. In general, any obtained increases were relatively small (4 to 9 percentage points) and may not merit consideration in light of the increased testing time necessary for administering multiple probes.

Summary and Considerations for Future Research

The purpose of this study was to examine the technical adequacy of three potential measures of algebra proficiency. One hundred thirteen students in grades eight to twelve participated in the study; 18 of these students were receiving special education services. The data were gathered in May, when students had nearly completed a year-long algebra course. On two occasions, students completed two forms of a Basic Skills probe, two forms of an Algebra Concepts probe, and a single form of a Content Analysis probe. The testing sessions were one week apart and were followed one week later by the administration of the Iowa Algebra Aptitude Test. Data collected on additional criterion variables included students' grades in school and in algebra, teachers' ratings of students' proficiency in algebra, percentages of homework completed and accuracy percentages for homework and tests/quizzes, and scores on standardized achievement tests. For the eighth graders in the sample, data were available for the math tests of the Iowa Tests of Basic Skills; for the ninth through twelfth grade students, the standardized achievement test data were drawn from the Iowa Tests of Educational Development.

In examining the distributions of student scores on each of the three types of measures, we found evidence of floor effects for the Basic Skills task, which was scored for the number of problems completed in one, two, and three minutes. Future research should examine the extension of time available to work on this measure to four minutes. In addition, we found evidence of a ceiling effect on the Algebra Concepts probe, which was scored for the number of problems completed correctly in five, six, and seven minutes. On this task, the mean scores across all participants were generally more than half the possible points on this task, even for the five minute timing. This finding suggests that using similar administration time limits might restrict the range for potential growth on these measures. Future research should examine the effects of shortening the time allowed to complete the probe (many students in our sample completed the entire probe prior to the end of the five minute time period), as well as examining whether similar ceiling effects occur when the probe is administered earlier in the algebra course. Finally, floor effects were also observed in the distributions for the Content Analysis probes. Despite an extended time period for administration, (10 minutes), many students (30 – 45%) obtained scores of 10 or less on this task. Future research is needed to determine whether this result will be replicated with other samples of students.

Our data on the reliability of the measures led us to conclude that the measures meet expected standards for reliability for research tasks. On the Basic Skills task, both alternate form and test-retest reliability levels were generally above .80 for single forms. No substantive gains were obtained when scores from two forms of the probe were aggregated. On the Algebra Concepts probe, reliability estimates ranged from .72 to .86 for single forms of the probe. Although moderate gains were achieved when two scores were aggregated (to .88 to .91), it is less clear whether the costs of doubling the administration time merit the gains in reliability. In part, practitioners will need to make this decision based on the context in which the data might be used. The Content Analysis probe had strong reliability estimates for single forms; no analyses of the effects of aggregation were possible because only a single form was administered during each data collection session.

With regard to criterion validity, our results showed clear distinctions between the measures. While the Basic Skills probe demonstrated strong relations with the computation subtests on the standardized tests, the correlations with other indicators were much lower (in the .3 to .5 range) or nonsignificant. If these findings are replicated in the two additional districts, we will conclude that the Basic Skills probe is not a good indicator of student proficiency in algebra. The Algebra Concepts probe showed moderate to strong relations with many of the criterion measures, with correlations ranging from .6 to .8. The correlations for the Content Analysis probe were only slightly lower (.5 to .7), but concerns about the limited range of performance must also be considered when evaluating the potential of this measure for use in progress monitoring. Our final set of analyses examined the extent to which scores on the three measures could discriminate between four groups of students (advanced, typical, slower pace, and special education). The Algebra Concepts task was the only measure on which all four groups of students demonstrated differences in mean performance that were statistically significant.

Future research involving the algebra progress monitoring measures should examine the following issues:

- The effects of extending the time available to complete the Basic Skills task on the distribution of scores and on the reliability and criterion validity of the measure
- The effects of reducing the time available to complete the Algebra Concepts task on the distribution of scores and on the reliability and criterion validity of the measure
- Possible revisions to the format, content, or timing of the Content Analysis probe if replication studies obtain similar findings
- The extent to which a ‘learning curve’ influences students’ scores on the probes. In other words, how many probes must be administered before students’ performance begins to stabilize and improvements can be attributed to learning the concepts and skills of algebra, rather than learning the nature and strategies of the task? This research will provide important data to inform decisions related to the use of the measures for screening and for progress monitoring.

References

Foegen, A. (2000). Technical adequacy of general outcome measures for middle school mathematics. *Diagnostique, 25*, 175-203.

Foegen, A., & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education, 35*, 4-16.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. *Diagnostique, 8*, 135-149.

Moses, R. P., & Cobb, C. E. (2002). *Radical equations: Civil rights from Mississippi to the Algebra Project*. Boston, MA: Beacon.

APPENDIX

Standardized Administration Directions

Basic Skills Probe – Form 1

Basic Skills Probe – Form 2

Algebra Concepts Probe – Form 1

Algebra Concepts Probe – Form 2

Content Analysis Probe – Form 1

Content Analysis Probe – Form 2

Sample Probe Directions

Algebra Probe Data Collection Procedures
XXX Jr./Sr. High
May 4, 2004

Teacher 1 Period 1

Materials: Probe Order: **A1, A2, B1, B2, C1**

1. Student copies of the probes
2. Stopwatch/timer
3. Pencils for students

General Introduction:

As you all know, your class and other algebra classes at Ballard are working with Iowa State on a research project to learn more about improving algebra teaching and learning. Today I need your help in trying out some of the brief tasks that teachers may be able to use to track student progress. As you may remember, ALL students will complete the tasks, but we will only use your scores in the research project if you and your parent or guardian have both given us permission to do so. Please clear your desk—the only thing you'll need for today's class is a pencil. (Distribute pencils to any students who need them.)

There are a few things you should know about the tasks, or probes, we will ask you to complete today. First, we will be limiting the amount of time you have to work on the tasks. We EXPECT that you will NOT be able to finish the probes. These tasks are different from classroom tests or quizzes and are not meant to be completely finished. Second, there may be problems on the probes that are difficult or unfamiliar. Please work across each row and consider each problem. If you do not know how to answer the question, put a small X in the box and continue to the next problem. DO NOT spend a great deal of time on any one problem. If you get to the end of the probe and still have time to work, go back to the problems you marked with an X and try to solve them. Third, for the first two types of probes we are trying to figure out the best amount of time to allow students to work on the probes. As you are taking these probes, I will say "SLASH" at two points. This means that you should draw an obvious slash mark after the problem you are working on. **(Demonstrate on board or overhead.)** Do you have any questions at this point?

Directions for Version A Probes: **A1, A2**

1. Distribute copies of the first Version A probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Say to the students:

    This is one type of task we are testing out. The problems on this probe include algebra equations using basic math facts, simplifying expressions by combining like terms, and using the distributive property to simplify expressions. Look at each problem carefully before you answer it. ***[DO NOT REPEAT THIS PARAGRAPH FOR THE SECOND ADMINISTRATION.]***

    When I say 'begin,' turn the sheet over and begin answering the problems. Start on the first problem on the left on the top row. Work across and then go the next row. If you can't answer the problem, make an 'X' on it and go on to the next one. Remember to make a slash mark when I say "slash." You will have 3 minutes to work.

3. Set timer for **3** minutes. Say ***Begin*** and start your stopwatch.
4. When timer reads **2** minutes, say ***Slash***.
5. When timer reads **1** minute, say ***Slash***.
6. When timer goes off, say ***Stop. Put your pencils down.***
7. Ask students to pass papers to the back of the room and prepare to repeat for second A version probe. Say, ***Now we will do another probe that is similar to the one you just finished.***

Directions for Version B Probes

1. Distribute copies of the first Version B probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Say to the students:

    This is another type of task we are testing out. The problems on this probe include translating words into expressions, solving simple equations, interpreting line graphs, and completing function or pattern tables. Look at each problem carefully before you answer it. ***[DO NOT REPEAT THIS PARAGRAPH FOR THE SECOND ADMINISTRATION.]***

    When I say 'begin,' turn the sheet over and begin answering the problems. Start on the first problem on the left on the top row. Work across and then go the next row. If you can't answer the problem, make an 'X' on it and go on to the next one. Remember to make a slash mark when I say "slash." You will have 7 minutes to work.

3. Set timer for **7** minutes. Say ***Begin*** and start your stopwatch.
4. When timer reads **2** minutes, say ***Slash***.
5. When timer reads **1** minute, say ***Slash***.
6. When timer goes off, say ***Stop. Put your pencils down.***

7. Ask students to pass papers to the back of the room and prepare to repeat for second B version probe. Say, ***Now we will do another probe that is similar to the one you just finished.***

<u>Directions for Version C Probes:</u> **C1**

1. Distribute copies of the Version C probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Say to the students:
3. ***3.*** This is a third type of task we are testing out. The problems on this probe represent the different types of problems that you are learning in your textbook. In general, you will probably find that the problems at the beginning are easier and those on the second page are more challenging. Look at each problem carefully before you answer it. The problems on this probe may seem more difficult than those on the probes you've already completed. If you would like to skip around as you answer the problems on this probe, you can do so; I will not ask you to make slashes during this probe.

4. ***When I say 'begin,' turn the sheet over and begin answering the problems. Start on the first problem on the left on the top row. Work across and then go the next row. If you can't answer the problem, make an 'X' on it and go on to the next one. Remember to make a slash mark when I say "slash." You will have 10 minutes to work.***

5. Set timer for **10** minutes. Say ***Begin*** and start your stopwatch.

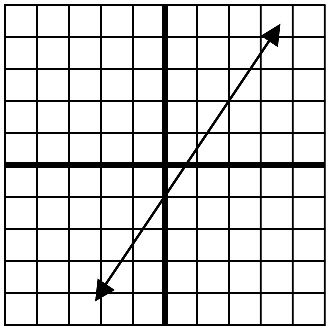
6. IF ANY STUDENTS COMPLETE ALL PROBLEMS DURING THE 10 MINUTE PERIOD, PLEASE NOTE THIS ON THIS PAGE!
7. When timer goes off, say ***Stop. Put your pencils down.***
8. Ask students to pass papers to the back of the room
9. Say, ***That is the end of the tasks for today. Next Tuesday we will be back in your class to do some more of the probes. Thank you for your help with our research project!***
10. If there is time left in the period, ask the teacher if s/he wants the students to do any particular activity OR play hangman with students until the bell rings.

Basic Skills Form 1

$9 + a = 15$ $a =$	Simplify: $2x + 4 + 3x + 5$	$12 - e = 4$ $e =$	Simplify: $(-5) + (-4) - 1$
Simplify: $4 - (-2) + 8$	$6 \bullet 9 = d$ $d =$	Simplify: $4(3 + s) - 7$	$2s = 16$ $s =$
$8m = 72$ $m =$	Simplify: $3(c + 2) - 2c$	$\frac{63}{c}$	
Simplify: $y^2 + y - 4y + 3y^2$	$y + 4 = 11$ $y =$	Simplify: $8m - 3(m - 2)$	$6 \bullet 3 = k$ $k =$
$12 - 6 = e$ $e =$	Simplify: $12 + (-8) + 3$	$\frac{z}{5} = 5$	
Simplify: $6r - 5 - 2r + 6$	$\frac{54}{w} = 9$		
$f - 7 = 3$ $f =$	Simplify: $-5 + 6 - 6$	$r - 4 = 4$ $r =$	Simplify: $w - w(4+5) - 6$
Simplify: $9 + (-3) - 8$	$\frac{36}{6} = s$		
$b + 8 = 8$ $b =$	Simplify: $4 - 7b + 5(b - 1)$	$\frac{18}{g} = 6$	
Simplify: $s + 2(s - 5) - 3$	$11 - n = 6$ $n =$	Simplify: $-3w^2 + 5w^2 - 5 + 12$	$f + 7 = 15$ $f =$
$4r = 28$ $r =$	Simplify: $5 - 2b + 4(b + 3)$	$2 + t = 7$ $t =$	Simplify: $1 - 9 + (-2)$
Simplify: $7b - 4 - 3 - 2b$	$5q = 30$ $q =$	Simplify: $3(u + 3) - 2u + 5$	$\frac{28}{4} = d$
$18 - 9 = k$ $k =$	Simplify: $-9 + 3 + 8$	$\frac{h}{6} = 8$	
Simplify: $16 + 2(t - 4) - 3t$	$4 + 7 = x$ $x =$	Simplify: $14 - 7 + (-3)$	$5 + z = 13$ $z =$
$5 \bullet 7 = j$ $j =$	Simplify: $9 - 4(v - 2)$	$12 - d = 4$ $d =$	Simplify: $b + b + 2b$

Basic Skills Form 2

$16 - p = 7$ $p =$	Simplify: $6 - 2(v - 7)$	$9 \bullet 5 = a$ $a =$	Simplify: $z + z + 3z$
Simplify: $15 + 3(y - 6) - 3y$	$8 + 3 = t$ $t =$	Simplify: $16 - 5 + (-3)$	$9 + b = 14$ $b =$
$12 - 5 = j$ $j =$	Simplify: $-7 + 11 + 2$	$\frac{h}{8} = 7$	
Simplify: $5q - 7 - 2 - 3q$	$8e = 40$ $e =$	Simplify: $5(u + 8) - 3u + 9$	$\frac{63}{9} = s$
$3t = 21$ $t =$	Simplify: $8 - 3g + 6(g + 2)$	$2 + a = 8$ $a =$	Simplify: $5 - 3 + (-8)$
Simplify: $p + 3(p - 6) - 4$	$13 - n = 5$ $n =$	Simplify: $-6m^2 + 2m^2 - 8 + 9$	$h + 8 = 11$ $h =$
$a + 5 = 5$ $a =$	Simplify: $6 - 9c + 7(c - 1)$	$\frac{21}{v} = 3$	
Simplify: $9 + (-4) - 8$	$\frac{49}{7} = w$		
$e - 5 = 4$ $e =$	Simplify: $-7 + 9 - 9$	$r - 2 = 6$ $r =$	Simplify: $f - f(2 + 8) - 7$
Simplify: $9k - 4 - 2k + 3$	$\frac{42}{s} = 6$		
$18 - 9 = p$ $p =$	Simplify: $13 + (- 7) + 5$	$\frac{d}{9} = 9$	
Simplify: $y^2 + y - 3y + 9y^2$	$x + 6 = 14$ $x =$	Simplify: $6c - 4(c - 7)$	$7 \bullet 3 = k$ $k =$
$9q = 81$ $q =$	Simplify: $4\ (c + 8) - 3c$	$\frac{56}{a} = 7$	
Simplify: $7 - (- 3) + 9$	$8 \bullet 6 = w$ $w =$	Simplify: $5(9 + s) - 4$	$2s = 18$ $s =$
$7 + b = 12$ $b =$	Simplify: $3m + 5 + 4m + 2$	$14 - e = 9$ $e =$	Simplify: $- 6 + (- 3) - 2$



What is the slope?

What is the y intercept?

Fill in the empty box:

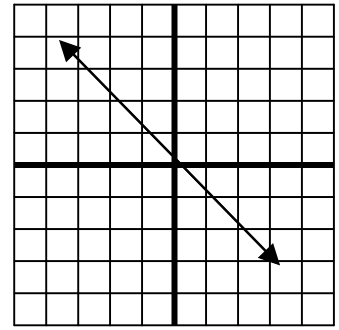
s	$3s$
6	18
7	21
8	
9	27

Fill in the empty box:

n	$4n+7$
1	11
2	15
3	
4	23

Fill in the empty box:

b	
5	2
3	0
0	-3
-2	-5



What is the slope?

What is the y intercept?

If $y > 9$, two possible values for y are _____ and _____

$6 \cdot 4 + 1 =$ ________

Simplify:
$7f + (2f + f)$

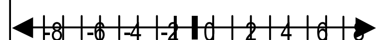
Solve:
$n + 3 = 8$
$n =$ ________

Evaluate $4b + 2$ when
$b = 1$ ________ and when
$b = 3$ ________

Write the expression for this phrase:
6 less than a number

$(-2) \cdot (-4) =$ ________

Graph the expression
$m > 6$



Write a word phrase for this expression:
$n + 9$

$8 \div 2 + 4 \cdot 3 =$ ________

$2^3 =$ ________

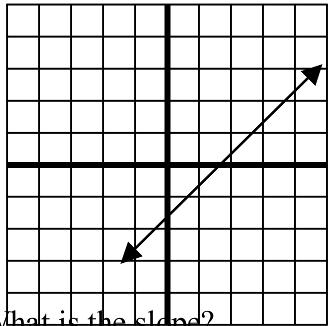
Write the expression for this phrase:
9 times a number

Write a word phrase for this expression:
$10b - 7$

Evaluate $2x + y$ when
$x = 2$ and $y = 3$
____

If $2a + 4 < 20$, two possible values for a are ________ and ________

Simplify:
$6 - 2(b - 4)$



What is the slope?

What is the y intercept?

Fill in the empty box:

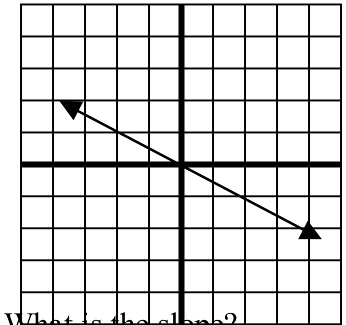
n	
6	4
9	6
12	8
15	10

Fill in the empty box:

t	t - 7
5	-2
6	-1
8	1
10	

Fill in the empty box:

w	
4	11
6	17
8	23
10	29



What is the slope?

What is the y intercept?

Write a word phrase for this expression:
$x \div 4$

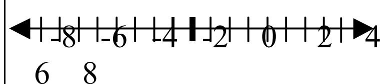
(-16) ÷ (-4) = _________

Write the expression for this phrase:
8 more than twice a number

Solve:
$3x = 27$
x = _________

Solve:
$6t = 36$
t =

Graph the expression $p \leq -3$



Simplify:
$9x - 3 - 4x + 9 =$

Solve:
$24 \div x = 6$
x = _________

Evaluate $8g - 4$ when
$g = 2$ _________
$g = 4$ _________

Write the expression for this phrase:
10 divided by a number

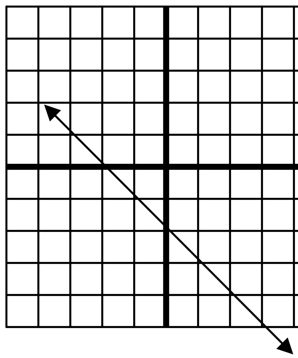
9 • 4 – 6 = _________

Simplify:
$12n - 5 - 7n + 3$

Write a word phrase for this expression:
4 times a number

(-3)(9 – 7) = _________

$\sqrt{\quad}$



What is the slope?

What is the y intercept?

Fill in the empty box:

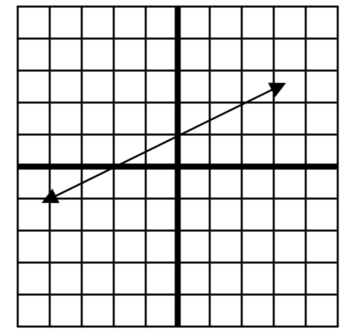
t	$4t + 1$
2	9
4	17
5	
10	41

Fill in the empty box:

y	
5	1
10	2
15	3
50	10

Fill in the empty box:

n	$n + 3$
1	4
2	5
3	
4	7



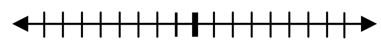
What is the slope?

What is the y intercept?

Write the expression for this phrase:
4 more than three times a number

Solve:
$49 \div n = 7$
$n =$

Graph the expression $t < 2$



$18 \div 3 + 6 \bullet 4 =$

Evaluate $6s + 4$ when
$s = 4$ ________ and when
$s = 6$ ________

Write a word phrase for this expression:
$c + 18$

$(-6) \bullet (-8) =$ ________

If $y > 3$, two possible values for y are ________ and ________

$(-5)(8 – 6) =$ ________

$\sqrt{49} =$ ________

Simplify:
$8t + (3t – t)$

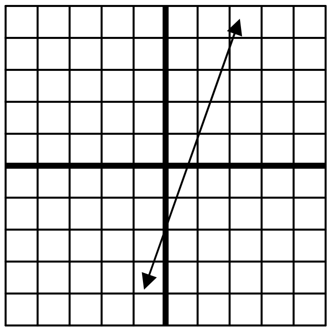
Solve:
$9x = 45$
$x =$ ________

$4 \bullet 5 – 2 + 6 =$ ________

If $2a + 4 \geq 12$, two possible values for a are ________ and ________

Write a word phrase for this expression:
$j – 12$

Simplify:
$2 + 2 \bullet 4 - 4$



What is the slope?

What is the y intercept?

Fill in the empty box:

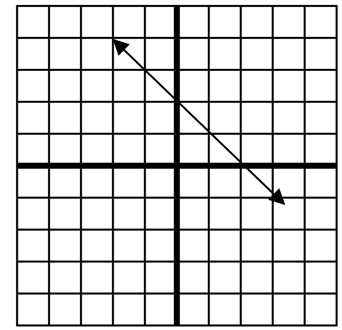
x	
10	5
20	10
50	25
100	50

Fill in the empty box:

n	$2n - 3$
3	3
5	7
7	11
9	

Fill in the empty box:

y	
1	8
2	9
5	12
10	17



What is the slope?

What is the y intercept?

Simplify:
$12n - 7 + 3n + 4$

Solve:
$b + 7 = 16$
$b =$

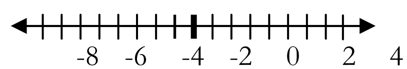
Write the expression for this phrase: *10 less than 3 times a number*

Write a word phrase for this expression:
$5t$

$-4(8 + 2) =$ _________

$3^3 =$ _________

Graph the expression $p \geq -5$



Evaluate $4a - b$ when
$a = 3$ and $b = 4$
____

Solve $18 - n = 12$
$n =$ _________

Simplify:
$9g + (2g - g)$

Write the expression for this phrase: *6 more than 4 times a number*

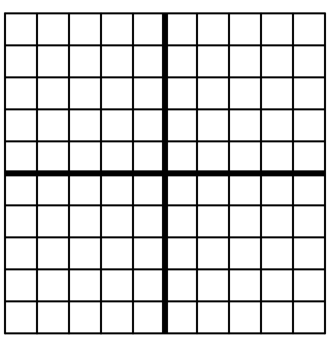
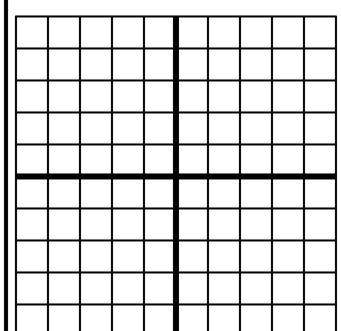
$8 \div 2 + 6 \bullet 2 =$

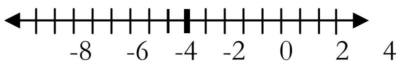
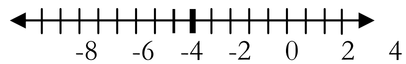
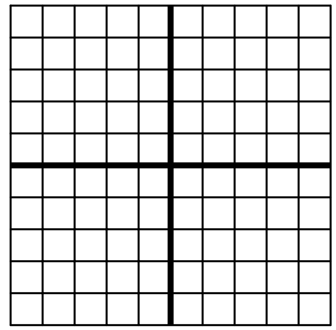
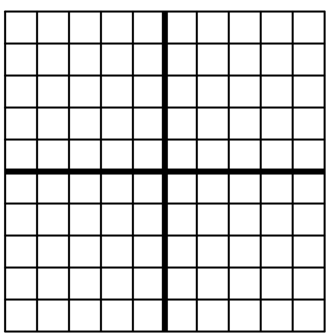
Write a word phrase for this expression:
$\frac{18}{b}$

Evaluate $5x - 4$ when
$x = 4$ _________
$x = 8$ _________

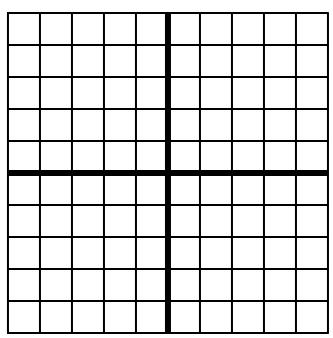
Write the expression for this phrase:
3 times a number

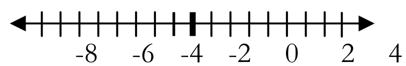
Simplify:
$5(m + 2) - 3m$

Evaluate $5y + x^2$ when $y = 3$ and $x = 2$	Simplify: $8(-4)(b)(b)(b)$	Solve: $\frac{x}{2} = (-6)$	Solve: $2t - 4 = 6$
Graph $x = 3$ 	Graph the equation: $y + 2x = 4$ 	Write the equation in slope-intercept form: $m =(-8)$ $b = 4$	Write the equation of a line that passes through (3, -1) and has a slope = -2. Use slope-intercept form.

Solve and graph: $x < (-2)$ or $3x - 5 > 1$  	Solve and graph: $(-5) + x \geq 1$  	Graph the linear system: $y = 2x - 3$ $-y = 2x - 1$ 	Simplify the expression: $\frac{x^3}{xy^4} \bullet \frac{y^5}{x^5}$
Circle the function that matches the graph: $y = \left(-x^2\right) - 2x + 3$ $y = \left(-3x^2\right) - x + 2$ $y = 2x^2 + x - 3$ 	Subtract: $(5t^2 - 9t + 1) - (8t + 13)$	Simplify the expression: $\frac{x^2 - x - 6}{x^2 - 4}$	Solve: $\sqrt{2m + 3} - 6 = 4$

Content Analysis Form 2, page 1

Add parentheses to make the expression true: $4 + 6 \div 2 = 5$	Rewrite without parentheses: $2(a + 3) - 2(a - 1)$	Solve: $6x - 9 = 10x + 3$	Solve: $5m = 25$
Find the slope of a line through (2, 4), (5, 0)	Graph the equation: $y = 2x$ 	Write the equation of a line through (-2, 5), (2, 4)	Write the equation in slope-intercept form: $m = 2$, b

Solve: $\lvert 2x \rvert > 14$	Solve and graph: $x < 2$ and $x \geq (-3)$ 	Solve the linear system: $(-6x) + 3y = (-6)$ $2x + 6y = 30$	Simplify the expression: $(mn)^2 \bullet n^4$
Evaluate the expression: $\sqrt{100}$	Solve: $(6x - 5)(x + 2) = 0$	Write the product in simplest form: $\frac{6x^2}{8x} \bullet \frac{-4x^3}{2x^2}$	Solve by completing the square: $x^2 - 2x = 2$